common methods for minimizing the effects of bias are described in Appendix I. If the study provides evidence that the investigators reduced the effects of bias, the methodologic quality grade was advanced to the next highest level.

It should be noted that inclusion criteria could influence report findings. The inclusion criteria chosen for this report permit review of the best evidence available on the clinical use of FDG PET scans for selected conditions. These generally represent larger controlled studies published in the peer-reviewed literature. A limitation of this analysis is the potential language bias owing to including only English language articles. Thus, the reader should keep in mind that the findings and recommendations are based only on evidence that meets criteria for inclusion in the report.

VII. APPRAISAL OF THE LITERATURE

For this update, titles and abstracts of 474 references were screened. Sixty-four references were determined to be relevant, and their full text articles were reviewed for potential inclusion in the systematic review. Additional articles were retrieved to provide background materials about the technology and selected clinical applications.

Forty-seven articles from the database searches and from end references of initially retrieved articles met the inclusion criteria for review. Each included study was classified according to clinical condition and assigned to a diagnostic efficacy level as follows:

Efficacy level*	Head & Neck	Breast	Lung staging	SPN	Colorectal	Alzheimer's
Technical	4	4	7	1	2	8
Diagnostic accuracy	3	6	7	2	3	0
Diagnostic thinking			?	?		
Therapeutic			?		?	
Patient outcome						
Societal						

* Adapted from Fryback and Thornbury, 1991
? Anecdotal data also presented in diagnostic accuracy studies.

In all oncology areas, higher levels of studies in the diagnostic test hierarchy superseded technical efficacy (feasibility) studies, represented the best evidence on the efficacy of FDG PET, and were summarized for this review. Technical efficacy studies are listed in the references. In Alzheimer's disease, only technical efficacy studies met the inclusion criteria for review.

All but one of the included studies were **single-site studies** classified as **case series** (Level V evidence), representing a relatively weak study design that does not provide strong evidence of effectiveness. Case series contain useful information about the clinical course and prognosis of patients, can *suggest* relationships between interventions and outcomes, and can generate ideas for further research. All studies used patients with no disease or with benign disease as **internal controls.**

All included studies used dedicated PET systems. The TA Program identified only one preliminary study using camera-based PET in oncology (Shreve, 1998). These authors compared blinded readings of camera-based PET images, using attenuation-corrected dedicated PET as the standard of reference, in 31 patients with known or suspected tumors. Accordingly, it did not meet criteria for inclusion in this review. The results are summarized below.

Table 5: Summary of the Technical Efficacy of Camera-based PET in 31 Patients with 109 Lesions

Site	*Short-axis diameter (cm) Range, mean*	*# lesions detected on camera-based PET*	*# lesions detected on dedicated PET*
Lung	0.9-4.0, 2.7	13	14
Mediastinum	0.6-1.3, 1.0	5	15
Mediastinum	1.5-3.5, 2.2	15	16
Axilla	1.2-1.5, 1.3	5	9
Head and neck	1.1-2.4, 1.7	5	7
Abdomen	1.2-6.3, 2.8	6	26
Skeleton	Not available, could not be determined	11	22

The authors concluded that camera-based FDG PET could depict many of the lesions depicted with dedicated PET. Detection of lesions using camera-based PET was greatest in the lung and poorest in the abdomen and in all sites, excluding the lungs, for tumors generally less than 1.5 cm in short-axis diameter. The results of this preliminary study require valid estimates of diagnostic accuracy and marginal value using an appropriate reference standard in order to establish camera-based PET as a diagnostic tool.

A. Data Synthesis

This report presents a qualitative overview to synthesize the best available evidence. A quantitative synthesis (meta-analysis) was not attempted. The methodological weaknesses of case series, combined with present differences in design and analysis among the eligible studies, argued against the validity and usefulness of pooling study results (Eysenck, 1994).

VIII. PUBLISHED FINDINGS

Background information on each clinical condition such as risk factors, diagnosis, alternative diagnostic modalities, staging, treatment and survival was described in detail in the first MDRC PET report (Flynn, 1996), and will not be presented here. A brief synopsis of updated epidemiological information and an account of the potential role(s) for PET are presented for each condition in addition to critical evaluation of the literature.

Epidemiological information for oncology conditions in this report is supplied by the American Cancer Society (American Cancer Society, 1998). Data on the veteran population are provided by the 1997 Annual Report of the Secretary of Veterans Affairs (West, 1998).

Results are presented according to the potential role of PET in the management of each disease. Full data abstraction tables of the best evidence of PET for each cancer section are found in Appendix IV.

A. Head and Neck Cancer

This report will define head and neck cancer as the common squamous cell carcinomas of the oral cavity, nasal cavity and paranasal sinuses, pharynx, and larynx. Skin, brain, thyroid, and salivary gland tumors and the rare tumors of other histopathologic types (sarcomas and lymphomas) that can have primary sites in the head and neck will not be discussed.

Approximately 41,400 new cases of head and neck cancer (3% of all incident cases of all types of cancer) and 12,300 deaths (2% of all cancer-related deaths) attributed to head and neck cancer are estimated for the United States in 1998. Within Veterans Health Administration malignant neoplasms of the lip, oral cavity, and pharynx (not larynx) accounted for 2,259 total discharges (0.3% of all discharges), with an average length of stay of 18.5 days, in FY 1997.

Nearly one-third of patients with head and neck cancer has lower stage, confined disease at diagnosis. Most of the remaining patients have locally or regionally advanced disease including spread to lymph nodes in the neck. Less frequent is head and neck cancer that has metastasized beyond the neck region (e.g., brain, lung, bone, or liver), at initial diagnosis. Accordingly, standard therapy emphasizes local and regional approaches (surgery, radiation therapy, or combination) with curative intent.

Chemotherapy is increasingly being added to standard therapy to improve the outcome of patients with locally advanced disease (PDQ®; 1999). For resectable disease neoadjuvant chemotherapy is incorporated into many organ preservation strategies to shrink tumors preoperatively and may improve locoregional control. Organ preservation approaches using concomitant chemotherapy with radiation are advocated in patients with unresectable disease.

Diagnostic tests are used at several points in the initial work up and treatment of head and neck cancer. These include delineating disease at the primary site (including locating unknown primary), identifying early nodal metastases, monitoring results of treatment, and identifying persistent and recurrent disease. CT and MRI have improved detection of occult cervical metastases for patients with head and neck cancer and subsequent management of patients at high risk of cervical metastases.

However, improvements are still needed to define the primary site and in the other points in the work up mentioned above. The ability to assess response to chemotherapy-radiation organ preservation approaches is becoming increasingly more important, since surgical excision would be indicated in the event of treatment failure. The functional information on glucose metabolism in head and neck tumors supplied by FDG PET could be clinically useful.

Table 6 depicts the study elements and Table 7 summarizes the data and quality of individual studies of PET using FDG in head and neck cancer.

Detecting unknown primaries in patients with metastatic cervical nodes

Braams (1997), a small technical feasibility study, detected unknown primaries in 13 patients with various histologic types of cervical metastases (see reference list). They performed whole-body PET followed by endoscopy, after physical exam and MRI and/or CT of the head and neck area failed to detect the primary tumor. PET identified the primary tumor in four (30%) patients and missed one small tumor (4mm) in another. Follow up over 18 to 30 months revealed no primary lesion in the remaining eight patients. The authors suggested that PET may be useful in guiding endoscopic exam and in identifying the primary site to direct more appropriate treatment.

Detecting primary disease

The MDRC Technology Assessment Program was unable to locate any PET studies that met evidenced-based criteria for diagnosing primary disease.

Detecting cervical node metastases

Two studies in Table 6 met some of the evidence-based medicine criteria for diagnostic test evaluations. Wong (1997) evaluated 16 patients, who had neck dissections, from a consecutive case series of 54 patients with known primary disease or with suspected recurrence or residual disease. Data suggest comparable performance of PET to anatomic imaging and improved performance over clinical exam across patients with a range of stages, but a test of statistical significance was not reported. In a small number of patients with occult nodal (N0) disease, PET did not perform as well as in patients with more advanced disease. In addition to small sample size in the subgroup analyses, several aspects of the study design were either unclear or not reported making the efficacy of PET difficult to determine.

In a retrospective evaluation of 14 patients with N0 disease on clinical exam, Myers (1998) reported a trend of increased accuracy of PET, although not statistically significant, over CT. PET combined with CT showed even greater improvement. Data were analyzed by dissected side and not by patient, and important study design elements were not reported.

Monitoring treatment response

Lowe (1997) presented preliminary data on 28 consecutive patients with advanced head and neck cancer, who were enrolled in a neoadjuvant organ-preservation protocol, to assess PET in evaluating tumor response to chemotherapy. The methods were reasonably well described, and the study met all evidence-based medicine criteria for diagnostic test evaluations. The data suggest good face accuracy of PET in distinguishing complete response from residual disease. Wide confidence intervals reflect a small study size, and no comparison data were presented.

The authors commented that while a positive PET scan may be indicative of residual tumor and warrant repeat tissue sampling or resection, a negative PET scan may also call for tissue sampling to rule out false negative results. They also stated that PET may be used in situations when sampling bias is more likely, for example, difficult access, questionable post-therapy biopsy results, or normal, reepithelialized appearance of the tumor site post-therapy.

Detecting recurrent disease

Wong and associates (1997) assessed PET prospectively for detecting both primary site recurrence in 12 patients and nodal recurrence in 13 post-treatment patients. PET showed high sensitivity in detecting recurrence at the primary site, but they presented no comparison data. For detecting nodal recurrence, PET was more sensitive than CT or MRI, was equal to clinical exam, and had superior specificity to both anatomic imaging and clinical exam.

Table 6: Characteristics of Diagnostic Accuracy Studies of FDG-PET in Patients with Head and Neck Cancer

Study characteristics	*Lowe et al. (1997)*	*Wong et al. (1997)*	*Myers et al. (1998)*
Perspective	? prospective	prospective	retrospective
Patient source	Consecutive patients between December 1994 and May 1996 with head and neck cancer: • 28 with stage III/IV who were participating in a neoadjuvant organ-preservation protocol using Taxol and carboplatin	54 consecutive patients who presented to head and neck clinics at two hospitals 31 with primary disease (TI=2 T2=10 T3=9 T4=10), 23 with suspected recurrence or residual disease) • 16 had neck dissections	116 patients diagnosed with head and neck cancer, of which 72 had biopsy-proven SCC and 26 underwent neck dissections: • 14 patients with N0 disease (24 total neck dissections) on clinical exam
Extent of disease (# patients)	Stage III=3 Stage IV =25	N0=8 N1=4 N2a=2 N2b=2	Stage I=1 Stage II=8 Stage III=2 Stage IV=3
Benign conditions	6 patients with pathologic complete response	None reported	None reported
PET criteria for positive result	1,2,or 3 on a 4 point scale	Not reported	Not reported
Contrast CT criteria for positive node	N/A	Standard size and morphological criteria used to assess nodal disease on CT/MRI	Not reported
Interpretation	• Blinded visual consensus using a before and after comparison format • 4-point scale • two readers	Not reported	Not reported
Gold standard determination (# patients)	Pathologic complete response or residual disease based on post therapy biopsies obtained after PET blinded to PET data (28)	• independent biopsy (16) • All suspicious areas of aerodigestive tract were biopsied	Histopathology for number of nodes, presence of malignancy, and extracapsular spread (14)
Data analysis	By patient	By patient	By dissection

Summary/Discussion

Since the 1996 MDRC PET report seven additional studies (three of diagnostic accuracy) of PET in head and neck cancer were published, met the inclusion criteria, and were reviewed. Evaluations of PET in head and neck cancer have focused mainly on detecting cervical node metastases in patients with known primaries, diagnosing disease recurrence, and monitoring response to treatment.

PET has potential uses at several points in the diagnosis and management of head and neck cancer patients. An early step in defining these uses is obtaining estimates of diagnostic accuracy. Only Lowe and associates (1997) met all evidence-based medicine criteria for diagnostic test evaluations, and the methods were reasonably well described. The two other studies did not report blinding of test interpreters and had other methodologic limitations, which affect the validity of the results, and it was unclear whether PET was used in addition to, or as a substitute for, other tests. All of the studies in Table 7 received low methodologic quality scores due to presence of significant bias, insufficient reporting and/or small sample sizes. The diagnostic accuracy estimates from these studies should be interpreted cautiously.

Information from a whole-body PET scan could have important treatment implications for patients with head and neck cancer. For example, identifying the primary tumor site not detected by other modalities could alter treatment planning. If the primary is from the head and neck, it is potentially curable with surgery and/or radiation therapy, whereas if the primary is located elsewhere, less toxic palliative treatment can be given. While there is a lower limit to the size of tumor that can be detected by PET, if validated in larger, rigorous studies, more accurate staging with PET could result in more appropriate treatment.

Minn et al (1997) (see reference list) assessed the feasibility of FDG uptake to predict cancer aggressiveness and survival. The results from 37 patients with primarily advanced Stage III/IV disease suggested a correlation between FDG uptake and prognostic significance on univariate analysis but not on multivariate analysis. Using FDG uptake to identify high-risk patients who would benefit from post-treatment surveillance requires further comparative study. Nonetheless, the wide range of primary sites and stages of head and neck cancer and the associated wide range of site-specific treatment and outcomes would complicate such evaluations of PET.

Accurate diagnosis of disease recurrence is critical to the treating clinician. With the addition of chemotherapy to many organ-sparing protocols, the ability to accurately assess nonsurgical treatment failure becomes increasingly more important to judicious surgical salvage. For patients who become symptomatic or who develop a mass during post-therapy surveillance, PET must be able to distinguish recurrence from treatment-related inflammation or fibrosis.

Goodwin (1998) suggested ways to improve such evaluations of PET that may provide more useful data to the treating physician. A prospective study of these patients, rather than a retrospective study of patients who had PET for various reasons and at various times after treatment, would more appropriately address the clinical issue. Pretreating patients with steroids or antibiotics to reduce inflammation might enhance the positive predictive value of PET. Other considerations include cost-effectiveness and capturing individual patient history, such as the timing of signs and symptoms after completion of therapy.

Controlled, prospective, blinded studies are needed to define the utility of PET (either dedicated or camera-based systems) relative to other imaging modalities in patients with head and neck cancer. Multiple sites may be needed to accrue a sufficient number of patients. Results from this updated literature review confirm the conclusions and recommendations from the first report (see Preface).

Table 7: Summary of Diagnostic Accuracy Studies of FDG-PET in Head and Neck Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*				*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET	*CT*	*MRI*	*Other*	*Comparison group*	*Gold standard*	*blinding*		
Detecting nodal metastases	Wong 1997	12 positive cases 4 negative cases	Se=67%	CT + MRI Se=67%		Clinical exam Se=58%	+	H	—	**S,R,W,t**	**D**
	Myers 1998	9 positive dissections 15 negative dissections	Se=78% Sp=100% PPV=100% NPV=88% Acc=92%**(P=0.11)	Se=57% Sp=90% PPV=80% NPV=75% Acc=76%*		PET + CT Se=86% Sp=100% PPV=100% NPV=91% Acc=95%	+	H	—	**S,r,W,t,d**	**D**
Detecting local recurrence	Wong 1997	10 positive cases 2 negative cases	Se=100%				+	H	—	**S,R,W,t**	**D**
	Lowe 1997	21 positive cases 6 negative cases	Se=90% (77-100%) Sp=83% (53-100%) PPV=95% NPV=71% Accuaracy=89%				+	H	+	**S,R**	**C**
Detecting nodal recurrence	Wong 1997	8 positive cases 5 negative cases	Se=100% Sp=100%	CT + MRI Se=75% Sp=80%		Clinical exam Se=100% Sp=60%	+	H	—	**S,R,W,t**	**D**

B. Breast cancer

The American Cancer Society estimates 180,300 new cases (178,700 women and 1,600 men) of breast cancer will be diagnosed in 1998 in the United States. After a 4% per year increase in the 1980s, breast cancer incidence rates have leveled off in recent years to about 110 cases per 100,000. An estimated 43,500 women and 400 men will die of breast cancer in 1998, making breast cancer the second major cause of cancer death in women. Mortality rates continue to decline, particularly in younger women, likely due to earlier detection and improved treatment.

In FY 1997, there were 1.2 million female veterans (4.8% of all veterans) living in the United States, and the percentage of females in the veteran population is expected to increase. In accordance with the Women Veterans Health Program Act of 1992, Health Services Research and Development supports research to increase outreach and access to health care and to explore health issues that affect many women, including breast cancer (Feussner, 1997). VHA has also established the Mammography Quality Standards Office and has made available a nationwide toll-free mammography information line (888-492-7844) to expand mammography services to female veterans.

Potential applications for PET in breast cancer management were defined previously (Flynn, 1996):

- Non-surgical evaluation of breast disease;
- Staging recurrent disease;
- Quantifying tumor glycolytic rate as a prognostic factor;
- Monitoring response to therapy;
- Patient selection for axillary dissection and for preoperative therapy;
- Screening in subgroups of women (eg, those with breast implants, with prior breast radiotherapy, multiple breast masses and history of negative biopsy results, or severely fibrocystic breasts).

Table 10 summarizes the data and quality of individual studies of PET using FDG in breast cancer. Only studies of dedicated PET for non-surgical diagnosis of breast disease, patient selection for axillary dissection, and staging recurrent/metastatic disease met the inclusion criteria for this review. Three studies evaluated quantitative indices of FDG uptake as an indicator of prognosis. These studies were classified as technical efficacy due to their preliminary nature and will be discussed in the Summary/Discussion section.

Defining unknown primary disease

Palmedo (1997) prospectively compared PET to scintimammography (SMM) using ^{99m}Tc MIBI in the pre-surgical evaluation of 20 patients with 22 suspicious primary lesions detected by clinical exam or mammography. The mean lesion size was 29mm (range 8-53mm), of which only 3 patients had lesions smaller than 9mm. Quantitative analysis of tracer uptake was also performed to characterize disease, but no cut-off value was defined prospectively. Anecdotal data suggested that PET was superior to SMM in detecting axillary lymph involvement, but

neither test could determine extent of disease. The authors stressed that the menstrual cycle and age, which can alter MIBI uptake and FDG uptake, respectively, in normal tissue and the methods used to calculate FDG uptake could affect test accuracy.

Detecting axillary lymph node involvement

The three studies in Table 8 met the inclusion criteria for review. Utech (1996), Crippa (1998), and Adler (1997) compared PET to axillary lymph node dissection (ALND) in patients with either suspected or confirmed breast cancer who were scheduled for axillary staging. Therapeutic decisions at surgery were based on clinical and routine imaging results, including mammography. PET was added in the test sequence after the routine work up as a potential noninvasive method for staging the axilla, the rationale being that a negative PET scan might obviate the need for ALND in selected patients and, thus, decrease the morbidity and costs associated with the procedure.

All were prospective studies, but only Crippa (1998) reported a consecutive series. The evidence for the use of PET in staging the axilla is confined to a select group of patients with a high prevalence of malignancy and few benign conditions. The extent of axillary disease, reported in two studies, was limited to patients with metastases to ipsilateral axillary nodes. Crippa (1998) provided limited evidence from small subgroups on the ability of PET to determine extent of disease, which is an important prognostic indicator; not surprisingly, PET sensitivity improved with more advanced disease.

Two studies used multiple readers to interpret PET images, but neither study assessed interobserver variability. Of note, Adler (1997) used a higher dose of tracer and longer scanning times than were used in other studies. All studies reported some evidence of blinding to the gold standard, but none met strict evidence-based criteria for blinding. Patient and disease characteristics, study design elements, and units of analysis varied across studies, and many study design elements were incompletely described or not reported, making the validity of these results difficult to assess.

Table 8: Characteristics of Prospective Studies of Axillary Lymph Node (N) Staging With FDG-PET in Patients with Potentially Operable Breast Cancer

Note: All studies included primary tumors of mixed histologies, primarily invasive ductal carcinoma.

Study Characteristics	*Utech et al. (1996)*	*Crippa et al. (1998)*	*Adler et al. (1997)*
Patient source	124 patients with newly diagnosed and histologically proven breast cancer prior to therapy • 64 patients with metastatic nodes • 60 w/ surgically negative axilla • ? consecutive series	68 consecutive patients (72 total axilla) with palpable breast nodules scheduled for surgery based on clinical and mammography/ultrasound results • 61 had ALND • no ALND in patients with benign lesions (8) and in situ ductal carcinoma (3)	From a larger prospective study of PET, 50 patients with 52 axillary dissections who met inclusion criteria: • age ≥ 30 years • ≥ 2 ALND within 3 mo. Of PET scan • ≥ 10 nodes dissected • ability to fast ≥ 4 hours • ?consecutive series
Exclusion criteria (# patients)	Hyperglycemic patients	None reported	• History of ipsilateral axillary lymph node dissection • Preoperative systemic therapy • Primary tumor < 5mm • Uninterpretable PET scan (2)
Benign conditions of breast (#patients)	None	• proliferative dysplasia without atypica (6) • focal inflammation (2)	None
Primary tumor size (mean, range)	Reported as: <1cm=16 >1cm=49 >2cm=30 >3cm=29	2.0 cm, 0.4-6.7cm	Reported as: T0=1 T1=31 T2=17 T3=3
Prevalence of confirmed N metastases (# positive patients/total patients)	44/124=35%	27/61=44%	20/52=38% (by axilla)
Extent of N metastases (# patients)	N0=79 N1=43 N2=2 • one with bilateral disease	N0=36 (# axilla) N1a=21 N1b=13 N2=2	Not reported
Axillary node size	Not reported	Not reported	Range <0.1cm-2.5cm
PET criteria for positive node	discrete focal uptake > background	focal uptake > surrounding tissue)	increased FDG uptake and scan quality; scores ≥ 3= positive on a 5-point scale
Interpretation	• 3 radiologists + 1 nuclear medicine • blinded to all data except primary tumor	• # readers not reported • blinded to histopathology, but to other information not reported	• two readers • independent, blinded to all but axilla side • discrepancies resolved by consensus
Gold standard determination (# patients)	• histology (104) • histology + follow up (20) • extensive nodal sampling (average #/patient=19, range 7-46)	• histology (61) • Extensive nodal sampling (average # /axilla=21, range 12-38)	• histology (50) • extensive nodal sampling (average #/patient=17, range not reported)
Data analysis	By patient	By axilla	By axilla

ALND=axillary lymph node dissection

Detecting recurrence and metastases

The two studies in Table 9 presented the best evidence on the use of PET to stage recurrent disease and metastases in breast cancer patients.

Table 9: Characteristics of Studies Using FDG PET to Stage Recurrent Disease and Metastases in Patients with Breast Cancer

Note: Both were retrospective studies.

Study Characteristics	*Bender et al. (1997)*	*Moon et al. (1998)*
Patient source	75 patients with suspected recurrent or with metastatic disease in undecided or equivocal cases • Includes results from CT/MRI • 63 patients had both PET and CT/MRI data available for comparison • ?consecutive series	57 female patients (83 lesion sites) with a clinical suspicion of recurrence not resolved by conventional imaging: • who underwent primary surgery with or without adjuvant chemo- or radiation therapy and • who were referred to the UCLA PET center from October 1990 to October 1995 • ?consecutive series
Exclusion criteria (# patients)	None reported	• patients who underwent chemo-or radiation therapy within 3 mo before PET • lesions that were biopsied • lesions diagnosed with known disease
Benign conditions of breast (#patients)	None	(# sites) • seroma (1) • muscle uptake (5) • thyroiditis (1) • radiation pneumonitis (1) • blood pool of great vessels (2) • osteoarthritis (1) • intestine (1) • unknown (6)
Primary tumor histology	Well-differentiated ductal carcinoma (46) Infiltrating lobular carcinoma (10)	Not reported
Prevalence of confirmed local recurrence (# patients)	14/63=22%	29/57=51%
Prevalence of confirmed N metastases (# positive patients/total patients)	17/63=27%	8/26=31% (reported by lesion site)
Extent of M metastases (# patients)	• Bone (15) • Lung (5) • Liver (2)	• Bone (16) • Lung/Chest wall (7) • Liver (2)
PET criteria for positive lesion	4 point qualitative scale (intense, moderate, low, none) • Positivity criteria not defined	5 point qualitative scale • scores $\geq$ 3=positive
CT/MRI criteria for positive lesion	not defined	N/A
Interpretation	• 2 readers • independent • not blinded to other data	• 3 readers, discrepancies resolved by 4^{h} reader • independent • blinded to histology but aware of suspicion of metastases
Gold standard determination (# patients)	• histology (71) • follow up (4)	• histology • lesion morphology on 2 or more conventional imaging studies • $\geq$ 6 months of clinical and radiographic follow up after PET
Data analysis	By patient	By patient and by lesion

Both studies were retrospective case series of patients with suspected recurrence and/or metastases and equivocal findings after conventional imaging. PET was used as a complement to conventional imaging. It was unclear whether the patients in these studies represented consecutive case series. It should be noted that Bender (1997) presented data on 75 patients, but only 63 patients had information on both PET and CT/MRI for direct comparison. Few benign

conditions were represented in either study. This may be an artifact of the work up, and the benign cases were likely identified prior to inclusion. Both studies had a higher proportion of patients with metastases to the bone than to lung and /or chest wall, or liver. It was difficult to compare other characteristics of the patient population across studies due to incomplete reporting or variations in the units of analysis.

Both studies used qualitative scales to define lesions on imaging and multiple readers to interpret the images. Moon (1998) presented some data on interobserver variability. Moon (1998) met most of the evidence-based medicine criteria for blinding, but Bender (1997) did not blind interpreters to other data.

Summary/Discussion

PET has several potential uses in the management of patients with breast cancer. Since 1996, four technical efficacy and six diagnostic accuracy efficacy studies were published that met inclusion criteria for the review, representing the best evidence supporting the use of PET in breast cancer management to date. No new studies were identified that assessed the role of PET in evaluating response to treatment or screening in subgroups of women, such as women with radiodense breasts or breast implants.

The evidence on the ability of PET to detect unknown primary disease for this report is limited to one small study comprising a select group with a high prevalence of malignancy and few patients with small primary lesions less than 1cm. Limitations in study design and reporting suggest the preliminary nature of this study. The results should be confirmed in a larger group of patients with a range of tumor sizes, benign conditions and stages of disease. Newer PET models with higher resolution and availability of new dedicated breast PET scanners may improve detection of smaller lesions (Wahl, 1998).

The current best evidence, derived exclusively from case series of patients with a high prevalence of malignancy and with few benign conditions, does not support the routine use of PET as the initial test in patient selection for ALND. At face value, the operating characteristics from these studies suggest that PET has a relatively high sensitivity with a lower positive predictive value and a correspondingly lower specificity with a higher negative predictive value as compared to ALND. PET also yielded a fair number of false positives, many of which could not be explained. Some of the more recent studies are larger, but methodologic biases and incomplete reporting justified low methodologic quality scores.

Variations in the characteristics of the study populations, scanning techniques, and in the units of analysis may affect the generalizability of these results, particularly to mammographically tested populations, which typically have a lower prevalence of malignancy. Predictive values and other estimates of diagnostic accuracy should be interpreted with caution.

ALND with histopathology of dissected nodes supplies critical information to treatment management, is currently recommended by the NCI for most patients with Stage 1 or higher disease, but is associated with significant morbidity. Relative to other studies of screening and treatment options, published PET data to date are based on small numbers of patients. Moreover, the lower boundary of resolution limits the ability of current PET modalities to detect tumors less than 1cm in diameter. The consequences of false negative PET results in the absence of ALND in patients for whom effective treatment is available should be avoided.

The potential for PET to visualize the internal mammary nodes (potentially N3 disease) has been reported (Wahl, 1998). An NCI-sponsored multi-center trial is evaluating the accuracy of PET in staging the axilla and will include patients with N3 disease (See Section IX). Clinicians should await the results of this study before incorporating PET into routine clinical practice.

Likewise, the evidence on use of PET in detecting recurrent disease and metastases and defining unknown breast disease is in its early stages. PET was typically part of a testing sequence, but the marginal value of PET in the work up of these patients remains to be determined. The authors emphasized, and the TA Program concurs with, the need for further studies to assess the clinical impact of PET in the management of recurrent breast cancer.

Utech (1996), Crippa (1998), and Oshida (1998) (See technical efficacy list in Reference Section) presented some evidence on the feasibility of using quantitative FDG PET uptake by either the primary tumor or axillary lymph nodes as a prognostic indicator. Any attempt to correlate PET data with survival requires knowledge of the underlying characteristics of the study population and sufficient follow up time to track survival (Laupacis, 1994). The range of disease stages and corresponding treatment options would further confound the results. Large, rigorous studies are needed to define the utility of PET as a prognostic test.

Controlled, prospective, blinded studies are needed to define the utility of PET (either dedicated or camera-based systems) relative to other imaging modalities in patients with breast cancer. Multiple sites may be needed to accrue a sufficient number of patients. Results from this updated literature review confirm the conclusions and recommendations from the first report (see Preface).

Table 10: Summary of Diagnostic Accuracy Studies of PET and Alternatives in Breast Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*			*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET	*SMM*	*CT/MRI*	*Comparison group*	*Gold standard*	*blinding*		
Defining unknown primary disease	Palmedo 1997	13 malignant lesions 7 benign lesions in 20 cases	Se=92% Sp=86%	Se=92% Sp=86%		+	H	+	**S,r,t,d**	**C**
Detecting axillary node involvement	Utech 1996	44 positive cases 80 negative cases	Se=100% Sp=75% PPV=69% NPV=100% Acc=84%			+	H + F	+	**r,d**	**C**
	Crippa 1998	27 positive axilla 45 negative axilla in 61 cases	Se=85% Sp=91% PPV=85% NPV=91% Acc=89% (overall values reported)			+	H	+	**S,r,t,d**	**C**
	Adler 1997	20 positive axilla 32 negative axilla in 50 cases	Se=95% Sp=66% PPV=63% NPV=95% Acc=77% (overall values)			+	H	+	**S,d,r**	**C**
Detecting recurrence or distant metastases	Bender 1997	54 positive cases 9 negative cases	Se=73-100% Sp=93-96% PPV=85-88% NPV=92-100% Acc=90-97%		Se=46-91% Sp=95-98% PPV=88-91% NPV=86-98% Acc=87-97%	+	H+F	—	**S,r,w,T,d**	**D**
	Moon 1998	29 positive cases 28 negative cases	Se=93% Sp=61-79% PPV=82% NPV=92% (overall values)			+	H+F	+	**S,R,t,d**	**C**

SMM = scintimammography with ^{99m}Tc-MIBI

C. Non-Small Cell Lung Cancer

Bronchogenic carcinoma, classified as either small cell or non-small cell, comprises 95% of all primary lung cancers. *This section will address only non-small cell varieties, as they constitute the majority (75%) of all bronchogenic carcinomas and, when localized, have the potential for cure with surgical resection.*

Bronchogenic carcinoma is the leading cause of cancer death in the United States. In 1998 the American Cancer Society estimates 171,500 new cases of lung cancer and 160,100 deaths from lung cancer. Malignant neoplasms of the bronchus and lung accounted for 9,730 discharges (1.5% of all discharges) with an average length of stay of 13.8 days within the Veterans Health Administration in FY 1997.

Non-small cell lung cancers (NSCLC) include adenocarcinoma (including bronchioalveolar), squamous (or epidermoid) cell carcinoma, and large cell (including large cell anaplastic) carcinoma. While 5-15% of NSCLCs are incidental findings on a chest x-ray, the vast majority of patients have symptomatic, advanced disease at clinical presentation.

Initial diagnosis is based on complete history, physical exam, and chest x-ray. If cancer is suspected, then staging is needed to assess the extent of local and distant disease. Stage of disease is the primary predictor of response to treatment and one of the important predictors of survival.

CT is the preferred diagnostic imaging test and is used at several points in the management of a patient with lung cancer: 1) to stage disease; 2) to evaluate treatment response; and 3) to differentiate recurrent disease from fibrosis. Use of other diagnostic imaging technologies to stage lung cancer is circumscribed largely because of technical limitations, availability, and cost.

CT provides morphologic (typically size) detail of the disease site. Accordingly, disease status of mediastinal lymph nodes are classified according to size, with nodes greater than 1 cm in diameter generally indicative of malignancy. This can be problematic, because benign lymph nodes may appear enlarged and micrometastases may appear normal on CT. Consequently, biopsy confirmation of the primary site and metastases is required to determine the most appropriate treatment.

More accurate noninvasive methods for staging NSCLC are needed to minimize the use of invasive procedures for diagnosis and monitoring treatment response. To this end, the metabolic information provided by a PET scan may be useful. Several roles for PET in staging lung cancer have been identified in the literature:

- Defining unknown primary disease;
- Detecting hilar and mediastinal metastases;
- Detecting distant metastases;
- Defining recurrence from fibrosis;
- Analyzing tumor biology;
- Monitoring response to therapy;

- Predicting tumor response by measuring uptake of chemotherapeutic agents.

Tables 11 and 12 depict study characteristics and Table 13 summarizes the data and quality of individual diagnostic accuracy studies of FDG-PET in NSCLC that met the inclusion criteria for this review. Scores were further refined with pluses and minuses to reflect the degree to which investigators minimized the effect of these biases on diagnostic accuracy results.

Defining unknown primary disease

Two studies met the inclusion criteria for the report. Guhlman (1997) and Hagberg (1997) are relatively small retrospective surgical series with a high prevalence of malignancy in their respective cohorts. Both evaluated PET in the test sequence after CT, but only Guhlman (1997) measured PET independently of other tests in all patients. Neither study presented data comparing PET to CT alone. Both studies received low methodologic quality grades due to incomplete reporting of methods and significant biases in study design, which may inflate estimates of diagnostic accuracy.

Detecting hilar/mediastinal adenopathy

Recent evidence on the use of PET in NSCLC emphasizes its staging potential. Six studies meeting the inclusion criteria presented evidence on the diagnostic accuracy of PET in nodal (N) staging and are listed in Table 13. All enrolled patients had suspected or biopsy-proven lung cancer. Data analyses included only biopsy-verified cases, implying a strong presence of work up bias across all studies. All studies assessed the role of PET independently of CT in the work up; Vansteenkiste (1997) also assessed PET as an adjunct to CT.

Guhlman (1997) and Hagberg (1997) were small retrospective studies with several methodologic flaws. The remaining four studies were reported as prospective evaluations of PET. Ambiguous descriptions of study methodology call into question the true, real-time prospective nature of three of them (Steinert, 1997; Vansteenkiste, 1997; Sasaki, 1996). Of these three, Sasaki (1996) was the most methodologically flawed.

Bury (1997) presented the largest and the only discernibly true prospective evaluation of PET in staging patients with NSCLC. Steinert (1997) and Vansteenkiste (1997) also presented notable attributes. These three studies represent the strongest evidence on the use of PET in N staging patients with NSCLC and are presented in Table 11 for comparison.

Table 11: Characteristics of Prospective Studies of Mediastinal Lymph Node (N) Staging With FDG-PET in Patients with Potentially Operable NSCLC

Note: All studies included mixed histologies, primarily squamous cell and adenocarcinoma.

Study Characteristics	*Bury et al. (1997)*	*Steinert et al. (1997)*	*Vansteenkiste et al. (1997)*
Patient source	141 consecutive patients who presented between 9/94-10/96 with new or suspected NSCLC based on sputum cytology, needle biopsy, or flexible bronchoscopy • 109 enrolled	62 surgical candidates with suspected or proven NSCLC who had PET between 2/94 and 3/96 • 47 enrolled	Unknown # patients who presented between 9/95-4/96 with suspected or confirmed NSCLC and who had standard M staging • 50 enrolled
Exclusion criteria (# patients)	• poor physiologic status (22) • poor compliance or no definitive diagnosis (11)	• prior neoadjuvant therapy • diabetes • inadequate CT (2) • distant metastases (8) • inadequate sampling (5)	• inoperable due to distant metastases • diabetes • treatment with oral corticosteroids • ischemic cardiomyopathy • direct mediastinal invasion of primary tumor • obvious bulky mediastinal adenopathies
Prevalence of confirmed N metastases (#N1-N3/# patients)	34/66=52%	29/47=62%	15/50=30%
Extent of N metastases (# patients)	N0=32 N1=20 N2=10 N3=4	N0=18 N1=16 N2=7 N3=6	N0=35 N2=15
Benign conditions	• nonspecific inflammation=2 • pneumonia=1 • multinodular goiter=1 • localized FDG uptake in hepatic-splenic angle of colon=1	none reported	none reported
PET criteria for positive node	• moderate uptake: > 2X uptake in contralateral or reference region • intense uptake: markedly higher than reference region	• FDG uptake ≥ FDG uptake in brain • nodular appearance	Grades 4 and 5 on a 5-point semiquantitative scale
Contrast CT criteria for positive node	short axis diameter > 10 mm	• short axis diameter > 10 mm except: • upper paratracheal nodes > 7mm short axis diameter • infracarinal station > 11 mm short axis diameter	maximal cross-sectional diameter ≥ 1.5 cm
Interpretation	• independent, blind • consensus by 2 radiologists and 2 nuclear medicine	• independent, blind • 1 radiology reader • 1 nuclear medicine reader	• independent, blind • one chest physician, one radiologist • 2 nuclear medicine readers
Gold standard determination (# patients)	• histology from mediastinoscopy (5), thoracotomy (51), both (10) • radiologic follow up based on CT or PET • all accessible nodes at surgery sampled	• extensive nodal sampling at thoracotomy of all identifiable nodes regardless of size on imaging • mediastinoscopy (22) and/or thoracotomy (18)	• nodal sampling at mediastinoscopy (47) and at thoracotomy (49), fine needle aspiration (1) • extent of sampling not reported
Data analysis	correlated by patient	correlated by nodal station	correlated by patient

Variations in study characteristics and units of analyses contributed to the range of reported estimates of diagnostic accuracy and differences in quality scores across studies. All studies had a significant degree of work up bias, which contributed to their low quality scores. All conducted varying degrees of nodal sampling, a means for minimizing diagnostic review bias, but the extent of sampling varied and was not reported with sufficient detail to enable the reader to quantify the effect of this bias on diagnostic accuracy. Bury (1997) and Vansteenkiste (1997) utilized multiple readers for blinded, independent image interpretation, but neither assessed interobserver variability.

Bury (1997) provided the strongest evidence to date on the diagnostic accuracy of PET in N staging NSCLC. A comparison of PET to CT yielded comparable accuracy estimates. The authors presented data on the impact of PET in modifying treatment, but no methods for systematic assessment were described. Bias in the stated methods and in incomplete reporting of other critical design elements hindered evaluation of study validity in the other studies. None of the studies assessed the incremental value of PET in the work up of NSCLC.

Detecting distant metastases

Studies in Table 12 met the inclusion criteria for review. Erasmus (1997) reported on 27 patients diagnosed with bronchogenic carcinoma and adrenal masses detected by CT. Adrenal masses are common in patients with NSCLC, but in the absence of other extrathoracic metastases, they are likely to be benign. Diagnosis of many adrenal masses remains indeterminate after standard anatomic imaging (CT or MRI), and a biopsy is required before treatment can be planned. The rationale for using PET in this case is to improve the noninvasive diagnostic accuracy, thus reducing the need for biopsy. Patients with normal FDG uptake in the adrenals and no evidence of distant metastases might be considered eligible for curative resection.

The findings suggest that, as an adjunct to CT, PET can discern malignant from benign adrenal masses using both visual and semiquantitative analyses. Results from this small preliminary study would need to be confirmed in larger, prospective studies to ascertain valid estimates of diagnostic accuracy and the added value of PET in diagnosing adrenal masses in these patients.

Bury (1997) present the strongest evidence to date on the use of PET for M staging NSCLC. They compared PET independently to conventional imaging (chest CT, abdominal CT, and bone scintigraphy) for M staging 109 patients with new or suspected NSCLC. The results suggest modest improvements in sensitivity and negative predictive value for PET over conventional imaging. The authors reported that PET correctly changed M stage, as determined by conventional imaging, in 14% of the cases and modified therapy in 20% of the patients, but the methods for assessing these changes were not described.

Table 12: Characteristics of Prospective Studies of Distant Metastases (M) Staging With FDG-PET in Patients with NSCLC

Study Characteristics	*Bury et al. (1997)*	*Erasmus et al. (1997)*
Patient source	141 consecutive patients with new or suspected NSCLC who had PET and conventional imaging between September 1994 and October 1996: • 109 patients enrolled in study • 39 patients with 59 sites of confirmed distant metastases	Unknown # consecutive cases presenting to thoracic surgery, oncology, or pulmonary between January 1993 and January 1996 with a diagnosis of bronchogenic carcinoma and an adrenal mass detected by CT • 27 patients with 33 adrenal masses enrolled in study
Exclusion criteria (# patients)	• Poor physiologic status (22) • Poor compliance or no definitive diagnosis (11)	• Inability to obtain informed consent • Poor clinical status • Death
Patient characteristics	• 77 men, 32 women • mean age= 64 yrs (44-83 yrs)	• 19 men, 8 women • mean age= 57 yrs. (39-76 yrs)
Characteristics of metastases (# patients)	• NSCLC (109) • Mean diameter not reported	• NSCLC (24); Small cell (3) • Bilateral masses (6) • Mean diameter=3 cm (1-9cm)
Prevalence of confirmed distant metastases	39 pts /109 pts=36%	23 sites /33 sites=70%
Locations of distant metastases (# sites)	• Adrenal glands(10) • Nonregional lymph nodes (6) • Lung (10); Bone (13); Liver (18) • Pleura (1); Soft tissue (1)	Adrenal glands (27)
Benign conditions (# sites)	• Nonspecific inflammation (2) • Pneumonia (1) • Multinodular goiter (1) • Localized FDG uptake in hepato-splenic angle of colon (1)	Not reported
PET criteria for positive metastases	• Moderate uptake: > 2X uptake in contralateral or reference region • intense uptake: markedly higher than reference region	Positive activity= activity > background
CT criteria for positive metastases	• Nodule characteristics not defined • Presence of clinical disease (symptomatic patient, progression on imaging, abnormal biochemistry) 6 months after imaging negative imaging	Visual detection of mass, characteristics not defined
Interpretation	• Independent, blinded to all data except histology of primary tumor • Consensus by 2 radiologists and 2 nuclear medicine	• Independent, blinded to clinical and biopsy findings • 3 readers
Gold standard determination (# patients or sites)	• Biopsy (21) • Clinical and radiologic follow up (88)	• Percutaneous needle biopsy (11) • Growth characteristics on sequential CT studies (16) • CT attenuation values < 10H (6)
Data analysis	Correlated by patient	correlated by site

Summary/Discussion

Early studies of PET suggested several potential uses for PET in managing NSCLC (Flynn, 1996). Positive trends in Medicare and private sector coverage policies for PET in lung cancer staging continue to fuel interest in the use of dedicated and camera-based PET as diagnostic tools. Since the first report, the TA Program identified 14 additional studies (7 of diagnostic accuracy) using dedicated PET, which met the inclusion criteria for this report. There were three areas in which potential uses for PET in NSCLC were studied: defining unknown primary disease, detecting nodal metastases, and detecting distant metastatic disease.

The best evidence on the diagnostic accuracy of PET in staging NSCLC suggests comparable accuracy of PET to CT in nodal staging and slightly better sensitivity, negative predictive value, and accuracy of PET over conventional imaging in staging distant metastases (Bury, 1997). Significant methodological biases, incomplete reporting of critical design elements, and variations in study characteristics (e.g., lack of uniform criteria for defining positive results on PET) limit the validity of the included studies and warranted low methodologic quality scores.

Appropriate use of the reference standard, or the "truth measure", is among the most challenging aspects of these studies to assess. Diagnostic review bias is often introduced, as biopsy sampling is rarely carried out independently of imaging results (e.g., it would be impractical to blind the surgeon to imaging). Bury (1997) minimized the effect of diagnostic review bias in nodal staging by conducting extensive nodal sampling and in distant staging by confirming disease status in all subjects using radiologic or clinical follow up or other confirmatory tests.

Imaging results are often used to determine which patients receive biopsy verification of mediastinal involvement (work up bias). To improve N staging accuracy several investigators advocated complementing the sensitivity of CT with the high negative predictive value of PET. They reasoned that a negative PET scan following a positive or indeterminate CT scan would exclude mediastinal metastases with a high degree of certainty and might obviate the need for invasive mediastinal evaluation (e.g., mediastinoscopy).

The best evidence for PET's N staging potential is confined to biopsy verified cases who had suspicious nodes on imaging. The size criteria for characterizing disease on CT and the lower detectable limit of resolution with PET may misclassify small tumor involvement, resulting in understaging. Failure to confirm disease status through follow-up in patients with negative CT or PET results may miss false negative results; failure to include the results in the analysis would result in inflated sensitivity and negative predictive values. Accurate, robust negative predictive values from studies that reduce the effect of work up bias are critical to determining the utility of PET in mediastinal staging.

Methodologically rigorous evaluations of diagnostic imaging, which reduced or accounted for the effects of methodologic biases on diagnostic accuracy, have been published (See Appendix II). In particular, Webb (1991) of the Radiologic Diagnostic Oncology Group (RDOG) provides an excellent model for evaluating diagnostic imaging in staging NSCLC. From patient enrollment to data analysis this rigorous evaluation offers extensive, detailed techniques for limiting the many biases inherent in diagnostic imaging studies. Incorporating study design elements from this model would strengthen the current best evidence for staging NSCLC using PET.

The value of diagnostic PET cannot be determined solely on improved accuracy over existing modalities. PET must demonstrate changes in diagnostic certainty and/or treatment planning or lower overall costs of patient management to justify its role in the work up. It can be argued that the metabolic information from PET may complement the information provided by conventional anatomic imaging and improve staging accuracy. More accurate staging may lead to more appropriate treatment planning. Studies included in this review reported anecdotal evidence of changes in treatment planning attributable to PET, but the impact of PET on treatment management was not systematically assessed, or reported as such. Furthermore, the range of stages and histologies of NSCLC and the associated range of treatments and outcomes would confound the effect of PET on outcomes of treatment, many of which are under investigation.

The TA Program concludes that the prevailing evidence does not support the routine use of either dedicated or camera-based PET in lung cancer staging. Data from rigorous, prospective clinical trials are needed to determine the added value of PET in the work up of NSCLC. Methodologically rigorous studies of diagnostic imaging have been published in the peer-reviewed literature. These studies may serve as models for guiding design of future PET research. Review of the more recent evidence confirms the conclusions from the first report.

Table 13: Summary of Diagnostic Accuracy Studies of PET and Alternatives in Staging Lung Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role *(Some assessed multiple roles)*	***Study***	***N***	***Operating Characteristics***			***Evidence-based Medicine Criteria***			***Study Design Limitations***	***Methodologic Quality Grade***
			PET alone (95%CI)	***PET + CT***	***CT alone (95%CI)***	***comparison group***	***gold standard***	***blinding***		
Defining unknown primary disease	Guhlman 1997	32 malignant cases 14 benign cases	Se=94% Sp=86% Acc=91%		Not reported	+	H	+	**S,r,w,d**	**C**
	Hagberg 1997	44 positive nodules 10 neg nodules (in 49 patients)		Se=93% Sp=70%	Not reported	+	H	—	**S,r,w,T,d**	**D**
Detecting mediastinal/hilar adenopathy	Bury 1997	34 positive cases 32 negative cases	Se=89% (72-96%) Sp=87% (71-97%) PPV=89% (72-96%) NPV=87% (71-96%) Acc=88%		Se=79% Sp=71% PPV=75% NPV=76% Acc=75%	+	H	+	**s,r,W,t,d**	**D+**
	Guhlman 1997	20 positive cases 12 negative cases	Se=80% (56-94%) Sp=100% (73-100%) Acc=87% (71-96%) (p<.02)		Se=50% (27-73%) Sp=75% (43-95%) Acc=59% (41-76%)	+	H	+	**S,R,W,d**	**D-**
	Hagberg 1997	9 positive nodes 9 negative nodes (in 18 patients with N2 disease only)	Se=67% Sp=100%		Se=56% Sp=100%	+	H	+	**S,r,W,d**	**D-**
	Steinert 1997	28 positive nodal stations 84 negative nodal stations (in 47 patients)	Se=89% (P=0.0066) Sp=99% PPV=96% NPV=97% Acc=97%		Se=57% Sp=94% PPV=76% NPV=87% Acc=85%	+	H	+	**S,r,W,d**	**D**
	Vansteenkiste 1997	15 positive cases 35 negative cases	Se=67% Sp=97% PPV=91% NPV=87% Acc=88%	Se=93% Sp=97% PPV=93% NPV=97% Acc=96%	Se=67% Sp=63% PPV=43% NPV=81% Acc=64%	+	H	+	**S,r,W,t,D**	**D**
	Sasaki 1996	17 positive regions 54 negative regions (in unknown # patients)	Se=76% Sp=98% (P<0.05) PPV=93% NPV=93% Acc=93% (P<0.05)		Se=65% Sp=87% (P<0.05) PPV=61% NPV=89% Acc=82% (P<0.05)	+	H	—	**S,R,W,T,d**	**D-**

Table 13 (cont.): Summary of Diagnostic Accuracy Studies of PET and Alternatives in Staging Lung Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role *(Some assessed multiple roles)*	***Study***	***N***	***Operating Characteristics***			***Evidence-based Medicine Criteria***			***Study Design Limitations***	***Methodologic Quality Grade***
			PET alone (95%CI)	***PET + CT***	***CT alone (95%CI)***	***comparison group***	***gold standard***	***blinding***		
Detecting distant metastases	Bury 1997	39 positive cases 70 negative cases	Se=100% (91-100%) Sp=94% (86-98%) PPV=90% (78-97%) NPV=100% (95-100%) Acc=96% (90-98%)		(conventional imaging*) Se=82% Sp=89% PPV=80% NPV=89% Acc=86%	+	H + F	+	**r,t,**	**B/C**
	Erasmus 1997	23 malignant lesions 10 benign lesions	Se=100% Sp=80%			+ internal	histology + CT follow up	+	**S,r,d**	**C**

* Bury et al 1997 defined conventional imaging as chest CT, abdominal CT, and bone scintigraphy

D. Solitary Pulmonary Nodules

Background information on solitary pulmonary nodules (SPN) is supplied by Lillington and Caskey (1993). A SPN is a single spherical lesion within the lung not associated with hilar enlargement or atelectasis and with a diameter generally less than 4.0 cm. The American Cancer Society reports that SPNs represent approximately 15% of all lung cancer diagnosed and estimates 25,725 new cases of malignant SPNs in the United States in 1998.

The differential diagnoses of a SPN include many malignant and benign processes. The most common malignant forms are bronchogenic carcinomas. Reported prevalence of malignant SPNs range from less than 5% to greater than 70% because of differences in the spectrum and severity of disease within each reported patient series. A malignant SPN represents a clinical stage I lesion, which is potentially curable with resection. Infectious granulomas represent the majority of benign processes and are caused predominately by coccidiomycosis, histoplasmosis, and tuberculosis.

The following risk factors directly correlate with the probability of cancer in patients with a SPN: 1) patient's age; 2) smoking history; 3) antecedent malignancy; 4) stability of lesion size on chest x-ray for 2 years; 5) absence of benign patterns of calcification within the nodule; and 6) nodule morphology (size and edge characteristics on CT). The baseline prevalence of malignancy in the study population may suggest the likelihood of a malignant SPN. Exposure to benign diseases such as tuberculosis or a history of residence in areas endemic for coccidiomycosis or histoplasmosis will suggest a lesser likelihood, but not rule out, malignancy.

Following clinical exam and chest radiography, the standard radiologic method of choice for evaluating SPNs is CT. CT provides information on the location and morphology of the nodule and can be used to guide biopsy procedures. Iodinated contrast material and high resolution CT densitometry may be used to enhance the differential diagnosis. However, limitations in the use of CT have been reported. Many SPNs are classified as "indeterminate" after CT and warrant invasive biopsy confirmation to determine the appropriate therapeutic course.

FDG PET has been proposed as a potential solution for improving the noninvasive differential diagnosis of SPNs, thereby reducing the need for higher risk invasive biopsy sampling and the associated morbidity and costs. Current evidence from this review supports the complementary use of PET after CT in the work up of patients with nodule diameters less than 3 cm or 4 cm, i.e., those nodules most likely to be indeterminate.

Table 14 displays the attributes of each study to highlight the variations in study quality and in criteria relevant to the applicability of the results. Table 15 summarizes the data and quality of individual diagnostic accuracy studies of FDG PET in SPNs.

Characterizing indeterminate solitary pulmonary nodules

Two studies met the inclusion criteria for this report. Dewan (1997) conducted a retrospective single-site study of indeterminate SPNs in 52 consecutive patients, who underwent PET between April 1990 and February 1994. They compared PET with and without standard criteria (clinical and radiologic data) using likelihood ratios[1] in Bayesian analysis to predict the probability of cancer in a SPN. Using sensitivity and specificity derived from this patient group, the authors determined that PET alone was the best predictor of cancer.

However, biases in study design and violation of the assumption of conditional independence between tests in the testing sequence, a requirement of Bayesian analysis, preclude drawing definitive conclusions regarding the accuracy of PET and its contribution to diagnostic certainty in these patients. Moreover, the impact of PET on treatment planning was not assessed. It is also important to note that many of these patients may have been included in studies assessed in the 1996 report.

Lowe (1998) conducted a multi-site study of radiologically indeterminate SPNs in 105 consecutive patients, who underwent imaging between October 1993 and August 1994. The study population included a broader range of benign conditions and nodule sizes compared with other published studies for this indication, reflecting the advantages of multi-site design. The authors presented a very detailed description of their blinding procedures and were the only investigators to calculate interobserver variability in visual analysis. From the stated methods, it is unclear whether they collected patient data in a "real-time" prospective fashion or retrospectively from surgical series.

These authors calculated likelihood ratios overall and for each subgroup. The likelihood of cancer was consistently higher using quantitative analysis over visual analysis. Except for specificity in SPNs ≤ 3cm in diameter, there were no significant differences between visual and quantitative analyses in the other diagnostic accuracy measures across subgroups. Small sample sizes in the subgroups likely contributed to the failure to detect any significant differences. Interobserver variability was very low (kappa=0.95), indicating good reproducibility of image interpretation.

[1] Likelihood ratio, expressed as Sensitivity/1-Specificity, is a measure of accuracy that indicates by how much a diagnostic test result will raise or lower the pretest probability of disease, thereby increasing the certainty about a positive or negative diagnosis.

Table 14: Characteristics of Studies Using FDG-PET of Patients with Radiographically Indeterminate Solitary Pulmonary Nodules

Study Characteristics	*Dewan et al. (1997)*	*Lowe et al. (1998)*
Perspective	Retrospective	Prospective (?not real-time)
Patient source	52 consecutive patients who underwent PET between April 1990 and February 1994 • included 3 with extrathoracic malignancy	Multisite study of 89 of 105 consecutive patients who underwent imaging between October 1993 and August 1994
Exclusion criteria (# patients)	• Cavitary or calcified nodules • Nodule size > 3cm • Age ≤ 30 years • # patients not reported	• no definitive histologic confirmation (8) • 4 not classified as radiographically indeterminate SPN (4) • no available CT scans (2) • nodule size < 0.7cm or > 4.0cm on CT(? # pts.)
Patient demographics	• 43 men (83%) • mean age ± SD=63.6±11.3 years • 41(79%) current smokers • 52% ≥ 20 cigs/day	• 61 men (69%) • mean age ± SD=63±9.5 years • smoking status not reported
Prevalence of malignancy	37/52=71%	60/89=67%
Nodule size in cm (%malig. pts. vs. %benign pts.)	≤ 1.0= 19% vs. 47% 1.1-2.0=51% vs. 40% 2.1-3.0=30% vs. 13%	0.7-1.5= 25% vs. 66% 1.6-3.0=60% vs. 24% 3.1-4.0=15% vs. 10%
Nodule Morphology (%malig. pts vs. % benign pts.)	Edge characteristics reported: • Sharp, smooth=14% vs.20% • Lobulated=30% vs. 40% • Slightly irregular w/ few spiculations=38% vs. 33% • Grossly irregular and spiculated=19% vs. 7%	Not reported
Benign conditions (#pts.)	• histoplasma granuloma with active inflammation (2) • other conditions not reported	• granuloma (7), coccidiomycosis (4), benign cellular debris (4), nonspecific inflammation (3), necrotizing granuloma (3) • fibrosis (1), hemangioma (1), aspergillosis (1), metaplasia (1)
PET criteria for positive node	focal FDG uptake > surrounding lung tissue, but more than mild intensity	• focal uptake > mediastinal blood pool structures (qualitative) • SUV> 2.5 (semiquantitative)
CT criteria for nodule edge	based on 4-type scale to reflect degree of spiculation and irregularity	not specified to image interpreters
Interpretation of PET	• qualitative • 1 reader blinded to histology • blinding to clinical and radiologic information varied	• semiquantitative using SUV • independent qualitative analysis using 2 readers blinded to clinical , imaging, and histopathologic data reached by consensus • readers interpreted studies with which they were not involved to ensure blinding • interobserver variability calculated
Interpretation of CT	• independent • 2 readers blinded to clinical diagnosis • consensus reading	• independent interpretation by > 1 reader blinded to clinical, PET, or histopathologic results • qualitative interpretation as benign or indeterminate
Gold standard determination (# patients)	thoracotomy (36), mediastinoscopy (3), bronchoscopy (3), needle lung biopsy (9), follow-up imaging for > 2 yrs (1)	TTNA (29) or surgery (60)
Data analysis	By patient	By patient

TTNA=Transthoracic Needle Aspiration

Summary/discussion

Since the 1996 report, three additional studies using dedicated PET in diagnosing solitary pulmonary nodules met the inclusion criteria for review. One was a technical feasibility study, and two were of diagnostic accuracy assessing PET in the test sequence after CT but prior to any histologic confirmation of disease. Both had significant biases in study design that warranted low methodologic quality scores and call for caution in generalizing these results to other populations.

Most false negative results reported in the PET literature are caused by small nodules with diameters commonly <1 cm that approach the resolution limits of the camera. Both studies reported false negatives comprising a variety of non-small cell cancers with diameters ranging from 1 cm to 2.5 cm. Moreover, the impact of PET on treatment planning, particularly the decision to proceed to surgery, was not systematically assessed.

One of the deficiencies outlined in the first report is the relatively low number of patients and a correspondingly narrow spectrum of benign conditions represented in the study base. Lowe (1998) presented the largest and only multi-site study of PET in diagnosing SPNs. Multi-site trials have the advantage of recruiting larger numbers of patients with a comprehensive array of malignant and benign conditions that are needed to apply the results to other populations. The detailed description of the blinding procedures used in the study may serve as a model for future studies of PET.

Both studies derived likelihood ratios (LR) to quantify the importance of the PET results in the work up of SPNs. As with predictive values, LRs are more useful accuracy measures to a clinician than sensitivity and specificity. LRs are used to calculate the probability of disease given a test result. They are independent of disease prevalence in most circumstances, but differences in case mix and methodologic biases can influence their validity (Gurney, 1993).

For example, the prevalence of malignancy in SPNs is lower in community hospitals than in most surgical series or in tertiary care facilities, where most PET scanners are found. Areas that experience a higher prevalence of particular benign conditions may encounter more false positive results on PET. A study with too few patients with benign nodules may overestimate specificity and inflate the negative LR; presence of methodologic biases may overestimate sensitivity and inflate the positive LR. In both studies the inclusion criteria favored a higher proportion of patients with malignancies and with too few benign conditions to offset the influence on specificity. Thus, rigorous study of a larger number and range of patients with a mix of diseases is needed to derive valid likelihood ratios for PET in patients with SPNs.

Table 15: Summary of the Diagnostic Accuracy and Diagnostic Thinking Efficacy Studies of PET in Indeterminate Solitary Pulmonary Nodules (SPN)

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*		*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET after CT Visual analysis (95%CI)	*PET after CT Semiquantitative analysis (95%CI)*	*comparison group*	*gold standard*	*blinding*		
Defining indeterminate SPN	Dewan 1997	32 malignant cases 14 benign cases	Se=95% Sp=87% Acc=92% LR_{mal}=7.11 (6.36-7.96) LR_{ben}=0.06 (0.05-0.07) *Overall values reported*		+	H + F	varied	**S,R,W,T,d**	**D**
	Lowe 1998	60 malignant cases 29 benign cases	Se=98% (95-100%) Sp=69% (57-81%) Acc=89% LR_{mal}=3.0 LR_{ben}=0.02 *Overall values reported*	Se=92% (82-100%) Sp=90% (79-100%) Acc=91% LR_{mal}=9.0 LR_{ben}=0.09 *Overall values reported*	+	H	+ (for visual analysis only)	**s,r,W,d**	**D**

Once valid LRs are derived, they may be used to estimate the odds that a patient has a cancer, given the PET result. Any attempt to use LRs in evaluating the odds of cancer after PET requires: 1) knowledge of the odds of cancer before PET, and 2) that the PET results were derived independent of the other test results. In neither study were both conditions satisfied, and the influence of PET on diagnostic certainty and subsequent treatment planning could not be determined.

Rigorous studies of patients comprising a range of pre-PET probabilities of malignancies are needed to assess the diagnostic accuracy and contribution of either dedicated or camera-based PET to the work up of solitary pulmonary nodules. Multiple sites may be needed to accrue a sufficient number and array of patients. Results from this review update confirm the conclusions and recommendations from the first report.

The Cooperative Studies Program of the VHA Office of Research and Development has funded a multi-year cooperative study to determine the efficacy of FDG-PET in defining solitary pulmonary nodules (See Section VIII). Results from this study should address the shortcomings of the existing literature.

E. Colorectal Cancer

Colorectal cancer is the third leading cause of death among men and women, representing a significant public health problem in the United States. Colorectal cancers account for approximately 11% of new cancer diagnoses. Death rates from colorectal cancer have fallen 25% for women and 13% for men during the past 20 years, reflecting a decreasing incidence of new cancer cases and increasing survival rates.

An estimated 131,600 cases and 56,500 deaths are attributable to colorectal cancer in the United States in 1998. An estimated 1 million veterans over the age of 50 will develop colorectal cancer over the remainder of their lives and nearly 433,000 will die from it (Wingo, 1995; Brown, 1996). Within the Veterans Health Administration, malignant neoplasms of the digestive organs and peritoneum (which include colorectal cancer) accounted for 8,280 discharges (1.2% of all discharges) with an average length of stay of 15.7 days in FY 1997.

Winawer (1997) reported the following risk factors for colorectal cancer: age over 50 years; a history of adenomatous polyps; a history of curative intent resection of colorectal cancer; inflammatory bowel disease; and familial colorectal cancer, adenomatous polyposis, or hereditary nonpolyposis colorectal cancer.

Nationally, the estimated relative five-year survival rate among veterans is approximately 40%, substantially lower than estimates from the general population of 62% (colon) and 59% (rectum). In VA, the Office of Research and Development (ORD)'s Epidemiologic Research and Information Center in Durham, North Carolina is conducting a four-year initiative to identify factors that may explain the worsened prognosis among veterans, and that may be responsive to intervention (Provenzale, 1998). ORD is also conducting a large prospective study of risk factors and/or detection of altered cell proliferation for

large colonic adenomas in asymptomatic subjects; the results will have important implications for colon cancer screening (Lieberman, 1998).

Data on management of colorectal cancer are from the National Cancer Institute's Physician Desk Query (PDQ) system retrieved in October 1998. The most prevalent histologic type of colorectal cancer is adenocarcinoma. Metastases to the liver, abdominal cavity, and extra-abdominal areas at initial diagnosis are common, as is recurrent disease after surgical resection of the primary tumor. Prognosis and management depends on the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases (staging).

Surgery is the primary therapy for colorectal cancer, and for cancers that have not metastasized, it is frequently curative. Many patients with confined recurrent disease or with metastases limited to the liver or lungs may also be amenable to resection. However, the high rate of recurrence and a troubling overall five-year survival rate call for more appropriate selection of patients who may benefit from surgical resection. The morbidity and costs associated with surgery for patients who do not have genuinely resectable recurrent tumor could be avoided by improved methods of tumor detection.

Stotland (1997) reviewed several imaging modalities commonly used to stage and diagnose colorectal cancer. The most common modalities include CT, MRI, endoscopic ultrasonography (EUS), and transabdominal ultrasonography. The popularity of EUS, in particular, has grown in recent years for its ability to image the depth of tumor penetration into the bowel wall and regional lymph node involvement. MR endorectal coils or ultrasound probes may be used to image rectal lesions. However, all structural imaging modalities are circumscribed in their ability to determine the presence and extent of disease and disease recurrence. Information from newer modalities, such as intraoperative ultrasonography, immunoscintigraphy, arterioportography, and PET, may increase the accuracy of staging and detecting recurrence.

Potential roles for PET in colorectal management have been identified in the literature:

- Pre-operative staging, including diagnosing presence and extent of liver metastases, and;
- Post-operative monitoring of recurrent disease.

Five studies met the inclusion criteria for review. Of these, two were technical efficacy studies and are listed in the reference section. Table 16 lists the characteristics of two retrospective case series and one prospective case series of diagnostic accuracy, and Table 17 summarizes the data and quality, representing the best evidence for the use of PET in managing patients with colorectal cancer. All studies presented some anecdotal evidence of therapeutic efficacy.

Preoperative staging of colorectal cancer

The TA Program identified one small uncontrolled, unblinded technical feasibility study of PET for staging initial *primary* colorectal cancer (Abdel-Nabi, 1998).

No diagnostic efficacy studies of staging primary colorectal carcinomas using PET were identified for review.

Four relatively small case series presented evidence on the use of PET in patients with suspected *recurrent* colorectal cancer, of which Ruhlmann (1997) was a retrospective technical feasibility study. The three remaining case series are diagnostic accuracy studies. Ogunbiyi (1997) and Flanagan (1998) are retrospective analyses from the same institution with overlapping study populations. Ogunbiyi (1997) studied 58 patients with a high suspicion for recurrence, including some with advanced primary disease, based on clinical symptoms, elevated plasma carcinoembryonic antigen (CEA) concentration, and/or CT findings. Flanagan (1998) assessed the ability of PET to detect recurrence in 22 asymptomatic patients with a post-operative elevated CEA concentration and normal clinical and radiologic findings.

Delbeke (1997) presented the only prospective comparison of PET to CT and CT arterial portography (CTAP) in detecting liver and extrahepatic metastases in 52 patients with suspected recurrent colorectal cancer. This is likely a continuation of an earlier, smaller study from the same institution (Vitola, 1996), which was reviewed in the previous 1996 MDRC technology assessment.

In all studies PET was performed as an adjunct to the routine clinical and radiologic work up, but the initial work up was not described in detail. Current evidence suggests that, when PET is added to the work up, there is improved sensitivity in distinguishing recurrence from post-surgical changes and documenting the presence and extent of liver and more distant metastases. However, the methodologic shortcomings in these studies limit the validity of these estimates. Predictive values may be subject to considerable referral bias owing to the high suspicion for malignancy in the study population. Lack of documentation of disease severity and underlying condition of the liver, completeness of the work up prior to PET, and blinding further hinders assessment of these results.

Table 16: Characteristics of Studies of Pre-operative Staging With FDG-PET in Patients with Suspected Recurrent Colorectal Cancer

Study Characteristics	*Delbeke et al. (1997)*	*Ogunbiyi et al. (1997)*	*Flanagan et al. (1998)*
Perspective	Prospective	Retrospective	Retrospective
Patient source	52 patients presented on 61 occasions with suspected recurrent carcinoma • Consecutive series	58 patients who had PET between 1/91 and 1/95 with suspected recurrent (n=47) or advanced primary (n=11) disease • ? Consecutive series	22 of 128 patients with history of colorectal cancer, who underwent PET from 6/93 to 6/96 • ? Consecutive series
Inclusion criteria	• Elevated CEA levels or abnormal CT • Abdominal CT (n=48); CTAP (n=40); or both (n=29)	• High clinical suspicion and equivocal or positive CT findings (n=39) • Elevated CEA levels with normal CT (n=19)	• Normal CEA levels after initial resection • Plasma CEA level > 5.0 ng/ml (mean 25 ng/ml), normal imaging studies, endoscopy, and physical exam on routine follow-up
Patient characteristics	• 31 men, 21 women • Mean age 63 ± 11 yrs	• 33 men, 25 women • Mean age 60 yrs. (23-81 yrs)	• 17 men, 5 women • Ages 17-84 • Primary site: colon (9), rectum (10), rectosigmoid (2), appendix (1)
Extent of disease (#patients)	• Liver metastases (45) • Extrahepatic disease (26, including 16 with liver mets)	• Primary disease or local recurrence (21) • Liver metastases (23) • Extrahepatic metastases (20)	• Stage B (10) • Stages C (5), C1 (2), C2 (3), • Stage D (2)
Benign conditions (# patients)	• Normal liver (7) • Post-surgical site (8) • Local fibrosis (2) • Resolving abscess (1), hepatic cyst (1), hematoma (1)	Not reported in reproducible detail	Not reported
PET criteria for positive site	• Not specified for qualitative PET • Cut-off not specified for semiquantitative analysis	• Malignancy=FDG uptake moderately or markedly intense; • Benign=no or mild uptake, or if abnormality identified on other imaging for which no corresponding abnormality was present on PET	Not specified
Contrast CT criteria for positive site	• Not specified for surgical cases • In nonsurgical cases, an increase in lesion volume > 20% on serial scans	Not specified	Not specified
CTAP criteria for positive site	• Not specified for surgical cases • In nonsurgical cases, an increase in lesion volume > 20% on serial scans	N/A	N/A
Interpretation	• 2 readers for PET , 2 readers for CT and CTAP, • Independent, qualitative PET blinded to other imaging results • Semiquantitative PET SUR calculations excluded lesions < 1 cm in diameter	• Qualitative PET interpreted with access to CT results • Two readers • CT interpreted in “routine clinical fashion”	• Qualitative PET scans interpreted with access to CT results • Consensus of at least two readers • CT interpreted in “routine clinical fashion”
Gold standard determination (# patients)	• Clinical or radiologic follow up (n=17) • Histopathology obtained surgically (n=44) • Percutaneous fine needle aspiration (n=2) • Nonresected lesions =surgical exam and intraoperative ultrasound (unknown #)	• Surgery, histology, or both (n=40); • Clinical and radiologic follow up (n=16); autopsy reports(n=2) • All patients followed for at least 12 months after PET or until death	• Pathology (n= 9) • All patients had radiologic and clinical follow up ≥ 6 months
Data analysis	By lesion site	By patient	By patient

Each study presented some evidence on changes in patient management attributable to PET, but the methods for assessment were not reported. The evidence suggests that adding PET to the work up may help optimize treatment (e.g., improve patient selection for curative surgery) by documenting the presence or absence of hepatic or more distant metastases. These data would need to be confirmed in much larger prospective studies designed to systematically assess the incremental value of PET against the many other available imaging modalities used in the work up of colorectal cancer.

Postoperative monitoring recurrent disease

The TA Program did not identify any studies in the published literature that addressed the role of PET in routine postoperative monitoring of patients for recurrent disease.

Summary/Discussion

Since the first report, five additional studies using dedicated PET in the management of colorectal cancer met the inclusion criteria for review. The best evidence to support the use of PET in colorectal cancers are three reported case series of diagnostic accuracy, of which two were retrospective studies from the same institution with overlapping study populations. All assessed the ability of PET as an adjunct to CT and other diagnostic tests to stage potentially operable patients with a high suspicion of recurrent disease; the one prospective case series also included patients with advanced primary disease. No diagnostic accuracy studies of PET to stage early, primary disease were identified.

Current evidence suggests that to further define recurrent disease, PET added after CT may offer improved sensitivity over CT alone. The absolute sensitivity of imaging modalities in detecting hepatic and more distant metastases is difficult to determine (Stark, 1987). ***Work-up bias*** is present when results from PET and/or other imaging tests under evaluation are used to direct biopsies to confirm suspicious liver lesions or to direct the choice of the most appropriate reference measure. Biopsy resection, while not entirely perfect, is a very accurate reference measure.

All authors attempted to offset work up bias by confirming disease in unresected patients using less perfect truth measures, such as clinical and radiologic follow-up, surgical exam and palpation, and intraoperative ultrasound. Although using these truth measures may not adequately identify the number of false negatives, they are reasonable alternatives and are preferred over nothing. The extent to which work up bias can be eliminated in this clinical setting is limited.

All of these studies had significant methodologic biases and insufficient reporting of fundamental design elements that preclude definitive assessment of study validity. The accuracy estimates from these studies should be interpreted with caution.

Table 17: Summary of Diagnostic Accuracy Studies of FDG-PET in Colorectal Cancer

H = histology; F = Follow-up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*			*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET after CT	*CTAP*	*CT alone*	*comparison group*	*gold standard*	*blinding*		
Diagnosing local recurrence	Ogunbiyi (1997)	21 recurrent cases 26 no recurrence	Se=90% (P=0.008) Sp=100% PPV=100% NPV=93% Acc=96%		Se=57% Sp=81% PPV=71% NPV=70% Acc=70%	+	Surg, H & F	—	**S,R,w,T,d**	**D**
	Flanagan (1998)	15 recurrent cases 7 no recurrences	Se=100% Sp=71% PPV=89% NPV=100%			+	H & F	—	**S,r,T,D**	**D**
Detecting liver metastases	Delbeke (1997)	104 malignant lesions 23 benign lesions in 45 patients	Se=91% Sp=95% Acc=92%	Se=97% Sp=5% Acc=80%	Se=81% Sp=60% Acc=78%	+	Surg, intra-operative US, H & F	partial	**S,r,W,t,d**	**D**
	Ogunbiyi (1997)	23 cases with disease 35 no disease	Se=96% (P=0.02) Sp=100% PPV=100% NPV=97% Acc=98%		Se=74% Sp=86% PPV=77% NPV=83% Acc=81%	+	Surg, H, F & autopsy	—	**S,R,w,T,d**	**D**
Detecting extrahepatic metastases	Delbeke et al. (1997)	34 malignant lesions 5 benign lesions in 26 patients	Se=100%		Se=74%	+	H & F	partial	**S,r,w,t,d**	**D**

CTAP=CT with arterial portography

All discussed changes in therapeutic management attributable to PET, but the methods for evaluation, details of the work up, or documentation of disease severity among the cases were not described. To suggest that PET improves the pre-operative staging process for selecting more appropriate patients for resection based on the existing evidence is ill-advised.

The TA Program did not identify any studies evaluating the efficacy of PET in post-operative monitoring. There is no consensus on the benefit of routine intensive follow-up after primary treatment, and the timing, frequency, type, and indications for post-operative follow-up using imaging are not standardized (Stotland, 1997). Any evaluation of PET in this role would be in the context of uncertain benefits of such monitoring.

Appendix II lists two particularly relevant studies for staging colorectal cancer and could serve as models for future PET research. Notable design features are highlighted. Zerhouni (1996) of the Radiology Diagnostic Oncology Group conducted a large, multi-site trial to compare the relative accuracies of CT and MRI in staging primary colorectal cancer. Stark (1987) compared CT and MRI to detect liver metastases, an important aspect of staging colorectal cancer patients. Studies of PET that incorporate these features with the comparable level of detail would provide more robust data on which to more confidently judge the added value of PET in the work up of colorectal cancer.

The TA Program concludes that the prevailing evidence does not support the routine use of either dedicated or camera-based PET in the management of colorectal cancer. Larger, prospective studies of diagnostic accuracy and subsequent therapeutic efficacy of PET in the work up are needed. Methodologically rigorous studies of diagnostic imaging have been published that may serve as models for guiding design of future PET research. Review of the recent evidence confirms the conclusions from the first report.

F. Alzheimer's Disease

This section briefly summarizes Alzheimer's disease (AD) and presents updated epidemiological information and results of a systematic review of the literature evaluating PET using FDG as a diagnostic test in AD. Appendix 8 of the MDRC technology assessment report on PET (Flynn, 1996) provides an expanded discussion of the disease, diagnosis, treatment, methodological and ethical considerations, and alternative neuroimaging technologies and other relevant diagnostic tests used in AD.

Unless otherwise noted, epidemiological information is from a consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society (Small, 1997). AD, a progressive neurodegenerative disorder, is the most common form of dementia and affects an estimated 4 million people in the United States. AD is characterized by steady irreversible decline in cognition, functioning, and behavior with sparing of motor and sensory functions until later stages. The rate of progression is variable, but duration of illness from diagnosis to death is approximately 10 years.

The reported prevalence of AD is approximately 6-8% of all persons 65 years or older. It doubles every 5 years after the age of 60 years, so that about 30% of the population older than 85 years will have AD. By the next century, an estimated 600,000 veterans with severe dementia will require long-term institutional care (ORD Impacts, 1997). The direct and indirect costs for care of AD patients in the United States approach $100 billion annually. The true costs of AD to society is likely much more, as economic assessments frequently underestimate the economic and emotional burden imposed on the caregivers as well as the patients.

Hendrie (1998) recently summarized the achievements in understanding genetic and nongenetic risk factors associated with AD. Genetic risk factors account for about 2% of all AD cases. Both causative (mutations on chromosomes 1, 12, 14, and 21) and associative genes (APOE-*4* allele[2] on chromosome 19) for AD have been identified. In VA, ORD researchers are: 1) studying genetic and environmental factors that contribute to delayed onset of AD in subjects with chromosome 1 mutations (ORD, 1997), and 2) are following subjects with the APOE-*4* allele at higher risk for developing AD to better detect and characterize early stages of this disease (Bondi, 1997).

Diagnostic tests that detect the presence of the APOE-*4* allele for apolipoprotein E, a serum lipoprotein involved in cholesterol transport, are under investigation, but experts differ on its usefulness. Since the APOE-*4* allele is found in many elderly persons without AD and is not always found in patients with AD, the Working Group of the American Medical Genetics/American Society of Human Genetics concluded that predictive testing of APOE-*4* for AD should not be done.

The only nongenetic risk factors consistently associated with risk for AD are age and family history. Other possible risk factors with a predominately positive association include low education, depression, estrogen-replacement therapy, nonsteroidal anti-inflammatory drugs (NSAIDs). Female gender, head injury, hypothyroidism and, to a lesser extent, insulin-dependent diabetes, aluminum exposure and smoking are inconsistently associated with an increased risk for AD. Clinical trials examining the role of estrogen, NSAIDs, and vitamin E in AD are reportedly underway.

The primary role of diagnostic testing is the differential diagnosis of AD from other reversible or treatable dementias. A definitive diagnosis is based on a typical clinical picture and histopathologic sampling of brain tissue at autopsy. In the absence of histologic confirmation, patients with probable AD are often referred to as having dementia of the Alzheimer's type (DAT). Two distinct sets of antemortem clinical criteria from the following may be used to characterize patients with DAT:

- (NINCDS/ADRDA)--National Institute of Neurologic and Communication Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
- (DSM-IIIR or the more recent DSM-IV)--Diagnostic and Statistical Manual for Mental Disorders, American Psychiatric Association.

[2] In Mendelian genetics, an allele is any alternative form of a gene at a given locus. An allele may express a dominant, a recessive, or an intermediate trait.

While advanced stage AD is usually easier to diagnose, early stage disease can be problematic. There is no cure for AD, but psychosocial techniques for behavioral problems associated with dementia and drug therapies for cognitive impairment have been developed, which can improve quality of life. HSR&D researchers found that two approaches improve quality of care and reduce costs associated with caring for AD patients: 1) simulated presence therapy, which uses selected memories through tape recorded conversations to manage problem behaviors in AD patients (Camberg, 1999); and 2) hospice care for managing AD patients with advanced dementia (Volicer, 1994).

New therapy aimed at slowing disease progression is also available. Since it is most effective if given at the earliest stages of AD, there is a need for obtaining earlier and more accurate antemortem diagnoses. Such information would also help patients and their families better prepare for future challenges. Functional imaging technologies such as PET and SPECT have been used to improve diagnostic certainty and to provide information on the pathophysiologic basis of AD.

Eight studies of technical efficacy using only dedicated PET scanners met the inclusion criteria for review. The TA Program was unable to identify published PET studies at higher levels of the Fryback and Thornbury diagnostic efficacy hierarchy. The following table summarizes information from these studies. All studies used FDG-PET to study regional cerebral glucose metabolic rates; Ishii (1997) also measured cerebellar glucose metabolic rates.

Evidence from recent technical efficacy studies shows a growing interest in the use of PET to better understand the biological mechanisms of neurodegenerative disease. The research suggests a link between cognitive function, functional imaging data, and the neurobiology of dementia. There is also increasing emphasis in these studies on improving methods for detecting early stage AD by improving the measurement of regional brain function. More precisely defined neuroanatomical atlases and methods of analysis may help explain the underlying pathophysiology of AD and the differences between diseases and disease progression.

Results from Imamura (1997) and Vander Borght (1997) underscore the limitations in existing knowledge using PET to diagnose AD. That is, while the temporal and parietal metabolic patterns often differentiate AD from other causes of dementia, AD also shares functional imaging features with other causes.

Table 18: Summary of Recent Technical Efficacy Studies Using FDG PET in Alzheimer's Disease

Study	*Objective*	*Findings suggest...*
Desgranges et al. (1998) N = 19	To study the neuronal basis for memory impairment in AD using Tulving's hierarchical model of memory systems and PET measurement of resting regional cerebral glucose utilization	• Their methodology for mapping neuronal substrates of cognitive impairment are valid and useful.
Higuchi et al. (1997) N = 20	To examine regional cerebral glucose metabolism using PET in AD patients with defined genetic risk factors (APOE-*4*, ACT, and PS-1 genotypes)	• APOE-*4* does not adversely affect the AD process or preserve brain metabolism after clinical onset of AD. • ACT gene has deleterious effects on cerebral glucose metabolism during the clinical stages of AD. • Differences in cerebral regions are influenced by the two genes. • Inheritance pattern of the two alleles may explain divergent patterns of progression in AD.
Imamura et al. (1997) N = 38	To study regional cerebral glucose metabolism in AD vs. dementia with Lewy bodies (DLB)	• There are differences in regional glucose hypometabolism consistent with the pathological and neurochemical differences between DLB and AD. • FDG-PET may help in the clinical discrimination between DLB and AD.
Ishii et al. (1997) N = 81	To study regional cerebral and cerebellar glucose metabolic rates in AD	• There is a significant cerebellar glucose metabolic reduction in severe AD with no apparent cerebellar atrophy. • AD is a global degenerative brain disease in which degeneration is correlated with severity. • Method of analysis using normalization of regional glucose metabolic data to cerebellar values may be liable to err in severe AD patients.
Pietrini et al. (1997) N = 16	To study regional glucose metabolism under stress using an audiovisual paradigm in nondemented adults with trisomy 21 Down's syndrome	• There are no differences in metabolism at rest. • In older subjects had significantly lower glucose metabolic rates in the parietal and temporal cortical areas. • A stress test paradigm can detect metabolic abnormalities in the preclinical stages of AD.
Stein et al. (1998) N = 50	Using a template of Brodmann areas derived from whole brain histological section atlas to analyze glucose metabolic rates in AD patients	• Vulnerability is greatest in cortical areas that are in closer synaptic contact with limbic areas. • Integrating statistical techniques of brain imaging into neuroanatomical atlases and incorporating fine-tuned calibration of neuroanatomical studies into brain-imaging analyses, may increase correlation of findings and a more complete characterization of the pathophysiology of AD.
Vander Borght et al. (1997) N = 27	To study regional cerebral glucose metabolism in AD vs. Parkinson's disease with dementia (PDD)	• AD and PDD may share common features in the patterns of metabolic alterations and also presence of regional metabolic differences in the visual cortex and in the medical temporal cortex. • These differences may help explain different degrees and combinations of disease specific underlying pathological and neurochemical processes.
Yamaguchi et al. (1997) N = 23	To study regional glucose metabolism in hippocampal atrophy in AD	• Morphologic asymmetry of the hippocampus and a metabolic asymmetry of the temporoparieto-occipital were correlated. • These asymmetries are present in early stage AD.

Summary/Discussion

Recent evidence exploits functional imaging technologies such as PET for pathophysiologic information that may be applied toward earlier preclinical diagnoses of AD. Jagust (1996) highlighted the importance and the complexities of obtaining earlier and more accurate diagnoses of AD:

- Earlier diagnosis is important for understanding the biological mechanisms of AD;

- Clinically, early diagnosis becomes more critical, as treatments become available;
- Information from early diagnoses may enable forecasting which elderly persons who experience memory lapses will develop dementia;
- Normal aging processes can complicate early diagnosis; and
- Research should also assess factors key to the production of disease symptoms.

The best evidence demonstrating the accuracy of FDG PET in diagnosing Alzheimer's disease is from four published studies reviewed by Flynn (1996). They are listed in the Alzheimer's disease references (Section XI). Although these studies reported good diagnostic accuracy for PET in AD, the diagnostic utility of PET remains controversial:

- While each set of clinical criteria has different associated sensitivity, specificity, and likelihood ratios, careful application of the clinical criteria does appear to identify most cases of treatable dementia.
- Sources of bias attributed to the spectrum and severity of disease, the use of clinical criteria as the gold standard, and the choice of clinical criteria (NINCDS/ADRDA versus DSM-IIIR or DSM-IV) may have influenced diagnostic accuracy estimates in these studies.
- Few studies applied PET prospectively to large numbers of patients with a spectrum of dementia and disease severity, which would be necessary to define the positive predictive value of PET as a diagnostic test, and followed them until death.

Flynn (1996) reported that a cooperative group of European PET centers is conducting such a study. The study will include patients with NINCDS/ADRDA "possible" AD, the patients in whom there is the greatest uncertainty regarding diagnosis and for whom a more accurate test would most contribute to posttest certainty.

Small (1997) suggested that improved diagnostic information to patients and their families may allow families to better prepare for the challenges ahead and that early and accurate diagnosis may prevent the use of costly medical resources. The TA Program was unable to locate any studies of PET that assessed the impact of PET on the costs associated with caring for patients with AD.

Flynn (1996) concluded that existing evidence argues against routine clinical use of PET for diagnosing AD until more effective treatments and risk modification interventions for AD are developed, and until meaningful and robust predictive values are obtained from an ongoing European multicenter PET study.